



# Developments in AQS Data Handling: Connecting the Dots

---

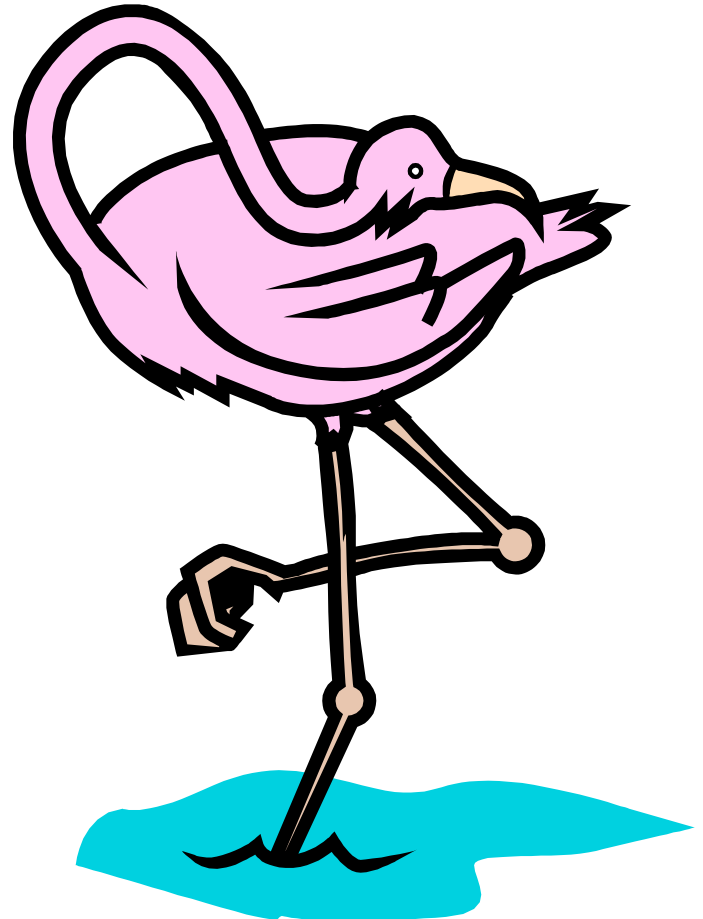
2004 AQS Conference  
Nick Mangus



# Three Big Topics

---

- Tribal Codes
  - Structure
  - Impact to AQS
  - Schedule
- XML Schema
  - And Web Services
  - Structure
  - Contents
  - Plans
- Data Access
  - Current options
  - New initiatives
    - Warehouses





# Tribal Site Codes in AQS

---

- Why
  - To honor tribal sovereignty
- How
  - Workgroup of tribal and federal members
  - Formed last year
  - Worked through several proposals
- What
  - A Way for tribes to enter data using codes entirely separate from state identifiers
  - AQS is one the first national system to make modifications to explicitly acknowledge tribal contributions
    - We're happy to be able to set an example



# Codes for Identification of Tribal Sites

---

- Reminder of how sites/monitors are identified in AQS
  - XX-YYY-ZZZZ
    - XX = "State" FIPS code
    - YYY = "County" FIPS code
    - ZZZZ = Site ID number (no inherent meaning)
- New way for tribal sites
  - XX = "TT" for all tribes
  - YYY = Tribal identifier code
    - Federal standard tribal codes (not census)
    - [www.epa.gov/edr/fdetribal.pdf](http://www.epa.gov/edr/fdetribal.pdf)
    - Alphanumeric
  - ZZZ = Site ID number as assigned by tribe



# Example Tribal Code

---

- Consider a Gila River site in Pinal County, Arizona
- Current way

"State" Code

04

"County" Code

021

Site ID

7701

- New Way

"State" Code

TT

"County" Code

614

Site ID

7701



# Impact on AQS Processes

---

- Site would have to be “pre-registered”
  - That is, contact your EPA regional office representative who will create the site with the ID’s you give them
- Data can then be loaded
- Retrievals
  - Criteria by tribe: just tribal ID’d monitors
  - Criteria by geography: all monitors in the area
- We’re close to finalizing the funding and the contractual issues
  - Will begin work soon
  - Hope to implement within one year



# Tribal Code Appendices

---

- This is **voluntary** – you do not have to do this for a site on tribal land
- Existing sites – if you want to change the ID, I think we can accommodate that (no unloading and reloading of all data)
- Remember state and county IDs are geographic, not political / ownership identifiers – you can currently find tribal data in AQS by selecting (in reports):
  - Reporting Organization (Supporting Agency Role)
  - Monitor Type (TRIBAL)
- When we have the chance (next time we re-engineer) we will try to be more correct in how we handle nations
  - Other non-U.S. data is currently short-cut
  - Cost prohibitive to incorporate now



# XML: What is it?

---

- eXtensible Markup Language
  - Adds data labels right into the data

## “Flat File” AQS

AA|I|34|054|0003|74.01354|34.5860|...

## “XML” AQS (anyone know HTML? SGML?)

<AirQualitySubmission>

  <NewSiteInsert>

    <TransactionDetails>

      <TransactionCode>AA</TransactionCode>

      <ActionCode>I</ActionCode>

      <StateCode>34</StateCode>

      <CountyCode>054</CountyCode>

      <SiteIDCode>0003</SiteIDCode> ... ..

      <LatitudeMeasure>-74.01354</LatitudeMeasure>

      <LongitudeMeasure>34.5860</LongitudeMeasure>





# How We Must Scheme

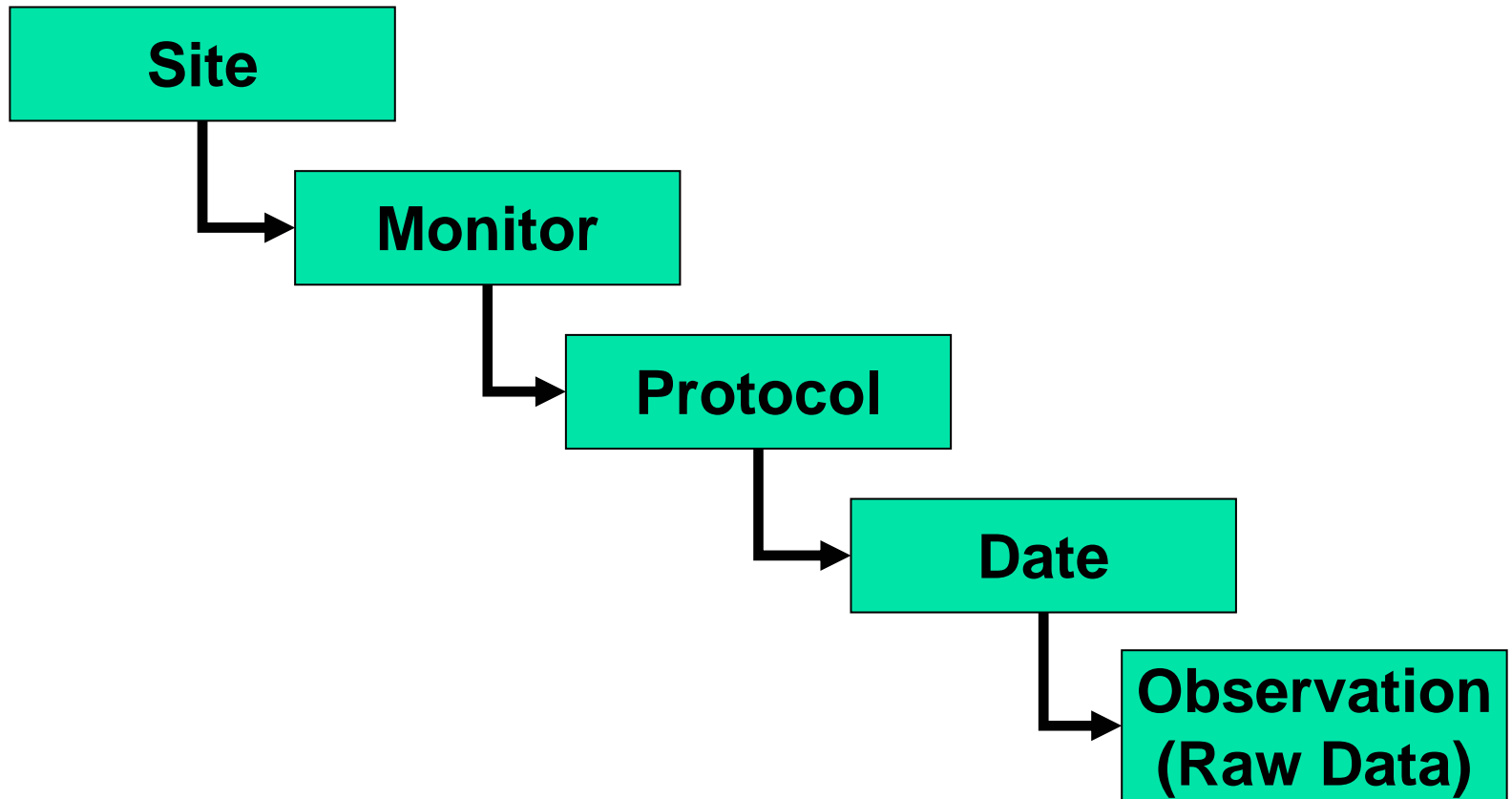
---

- A “schema” is what defines the allowable labels
  - Data combinations / hierarchies
  - Formats, allowable values, etc.
- A schema is analogous to a paper form that holds the data
- The data is the data – what you fill in the form with
- Ours is based on the transaction formats
- With hierarchical structure



# Structure of AQS Part of Schema

---





## AirQualitySubmissionType

### AirQualityNewSiteInsert

0..∞  
New site insert, site transactions  
AA-AC as needed

### AirQualityExistingSiteInsert

0..∞  
Existing site insert, site transactions  
AB-AC as needed

### AirQualitySiteUpdate

0..∞  
Existing site update, site  
transactions AA-AC as needed

### AirQualityEntireSiteDelete

0..∞  
Entire site delete, site transaction AA

### AirQualityPartialSiteDelete

0..∞  
Partial site delete, site transactions  
AB-AC as needed

### AirQualityNewMonitorInsert

0..∞  
New monitor insert, monitor transactions  
MA-MK as needed

### AirQualityExistingMonitorInsert

0..∞  
Existing monitor insert, monitor transactions  
MB-MK as needed

### AirQualityMonitorUpdate

0..∞  
Existing monitor update, monitor  
transactions MA-MK as needed

### AirQualityEntireMonitorDelete

0..∞  
Entire monitor delete, monitor transaction  
MA

### AirQualityPartialMonitorDelete

0..∞  
Partial monitor delete, monitor transactions  
MB-MK as needed

### AirQualityRawDataInsert

0..∞  
Raw data insert, transactions RC,  
RD, RA, RP, RS as needed

### AirQualityRawDataUpdate

0..∞  
Raw data update, transactions RC,  
RD, RA, RP, RS as needed

### AirQualityRawDataDelete

0..∞  
Raw data delete, transactions RC,  
RD, RP, RS as needed

### AirQualitySubmission

Top-level element for all AQS  
submissions

## AirQualityNewSiteInsertType

### TransactionIdentifierDetails

Required elements for all transactions

### TransactionAAInsertDetails

Basic site information action I details

### TransactionABInsertDetails

0..∞  
Site tangent street action I details

### TransactionACInsertDetails

0..∞  
Site open path action I details



### AirQualityRawDataInsert

Raw data insert, transactions RC,  
RD, RA, RP, RS as needed

## AirQualityRawDataInsertType

### TransactionIdentifierDetails

Required elements for all transactions

### TransactionParameterDetails

Required elements for all monitor  
transactions

### TransactionProtocolDetails

Protocol elements for select transactions

### TransactionDateDetails

0..∞  
Date elements for raw data  
transactions

### TransactionRCInsertDetails

0..∞  
Composite raw data action I details

### TransactionRDInsertDetails

0..∞  
Hourly, daily and sub hourly raw data  
action I details

### TransactionRAInsertDetails

0..∞  
Accuracy data action I details

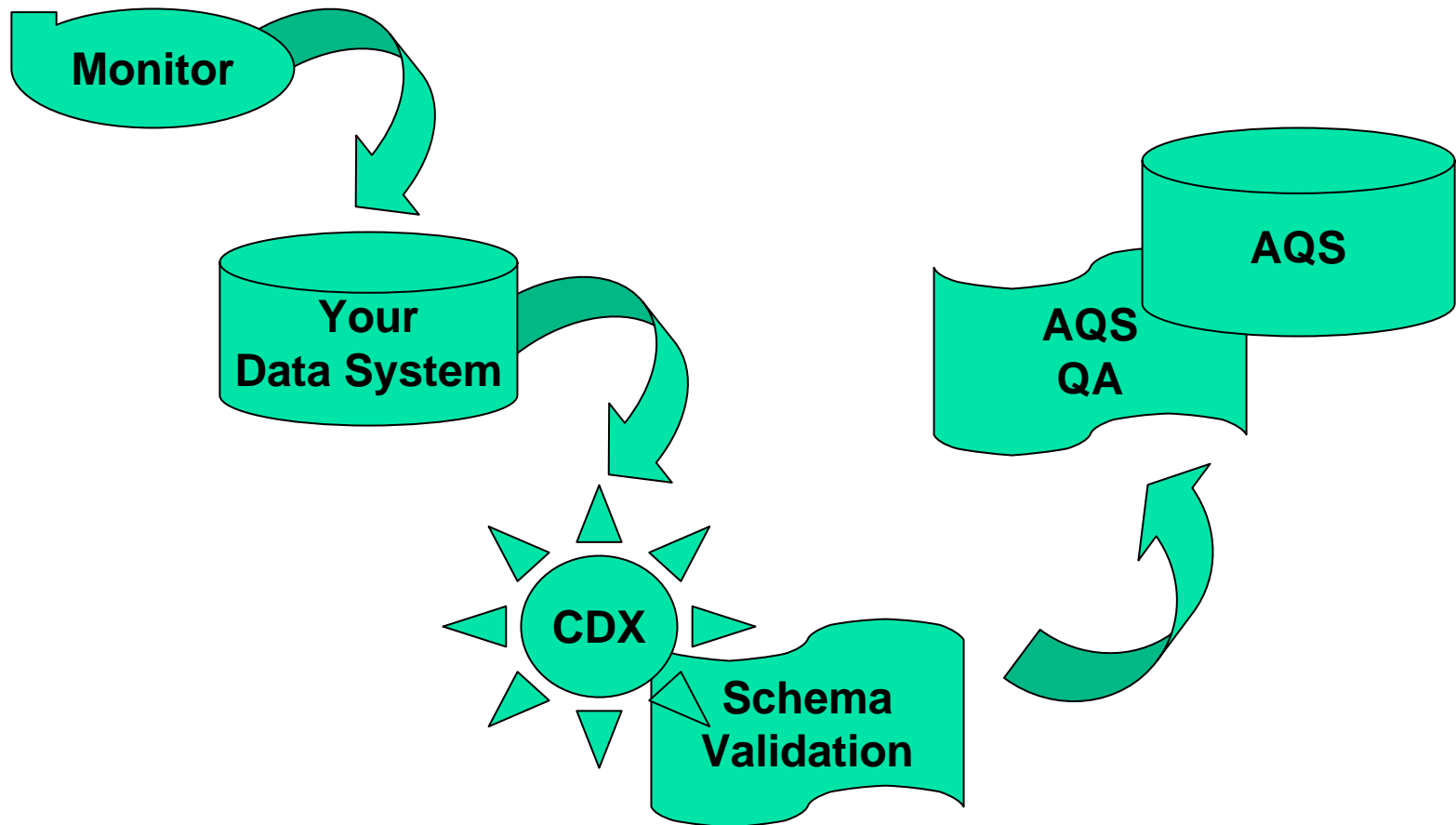
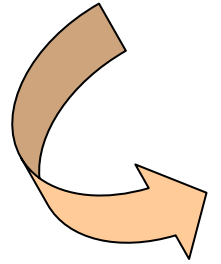
### TransactionRPInsertDetails

0..∞  
Precision data action I details

### TransactionRSInsertDetails

0..∞  
Annual summary data action I details

# XML Data Flow into AQS





# What's In It For Me, Nick?

---

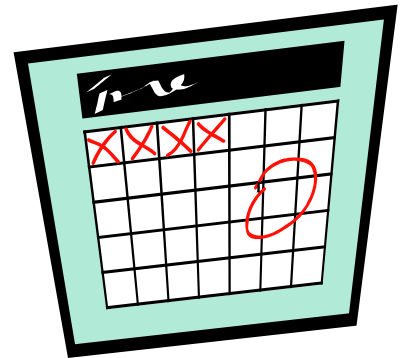
- We're planning on adding necessary elements for AirNow to schema ("Air Quality")
- Then for AQS output fields
- Allows for universal transfer of this data, not just you to us
  - Back and forth to each other
  - From the logger to you
  - To your IT department for public access
  - Most spreadsheets and DBs now compatible
  - Necessary step for web services (q.v.)
- Makes the data self-describing to other applications / users



# Schema Plans

---

- We're in the final draft for the AQS part of the schema
- Will finalize version 1.0 in late June
- Will modify AQS and CDX to accept XML files by fall
- Will be able to validate XML files at exchange network mid-summer
- Note: XML files are bigger than flat files
  - “|” replaced with tags (<value></value>)
  - 2.25 ratio file size increase for AQS raw data
  - Much better than 8:1 that others experience
  - Room for more improvement?





# Feedback

---

- Available at [exchangenetwork.net](http://exchangenetwork.net)
- We'd like your feedback
- By June 23
- E-mail [mangus.nick@epa.gov](mailto:mangus.nick@epa.gov)
- Helpful tools
  - XML software (commercial or free: IBM Xeena)
  - Can look at in web browser



# Web Services for AQS

---

- XML is a key component of web services
  - Necessary but not sufficient piece
- HTML is used for machines to make information available to humans
  - Via web browsers
- XML is used for machines to make information available to other machines (software)
  - Via web services
  - AQS
    - QA
    - Reports (q.v.)





# AQS Reports via Web Services

---

- How this will work:
  - Your application (e.g., spreadsheet) can download the service description
  - The description tells you where and how to send a request (exchangene트워크.net)
  - And what information to send with the request (geography, time, parameter, etc.)
  - Make the request and the report data is returned in XML
- By fall we plan to have several (2-4) AQS reports available
  - Popular but not computationally intensive
    - Quick Look
    - Data Completeness
    - Air Quality Index Summary
    - Maximum Values



# Getting Data from AQS

---

- Current means
  - Run AQS reports
    - Only available to registered users (~630)
  - Call EPA (HQ or RO) to request that we run a report
    - Or use the new web form on the TTN
  - Raw Data page on TTN
  - AIRData web pages (including AQS Query)
- Future Plan: Add Data Warehouse / Mart




## Technology Transfer Network Air Quality System

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) Search:  [GO](#)

[EPA Home](#) > [TTN](#) > [AQS](#) > [Requesting AQS Data](#) > Download Detailed AQS Data

### Download Detailed AQS Data

The table below provides links to files of detail data (a.k.a., raw data) extracted from the AQS database in [input transaction format](#) (RD transactions). These files following the naming convention of "RD\_501\_ppppp\_yyyy.ZIP", where "ppppp" = AQS parameter code and "yyyy" = year of data. All files are zipped to conserve disk space. Refer to our [Tools](#) page if you need assistance with file types.

A file containing 2000-2003 [site and monitor records](#) as of 3/5/04 in [input transaction format](#)  is also available as a text file. Site records begin with the letter "A"; monitor records begin with "M".

### Raw Data retrieved from AQS

Pollutant (parameter code) File name (substitute year for "yyyy")	Year of data, zipped file size, (unzipped file sizes), Date of extraction									
	Updates for these years may come at any time					Updates for these years are no longer expected				
	2003 most are incomplete until at least 7/1/04	2002	2001	2000	1999	1998	1997	1996	1995	1994
Carbon Monoxide (42101) RD_501_42401_yyyy.ZIP	<a href="#">42101_2003</a> 13711KB (241645KB) 3/20/04	<a href="#">42101_2002</a> 14585KB (261659KB) 3/16/04	<a href="#">42101_2001</a> 14581KB (269358KB) 3/19/04	<a href="#">42101_2000</a> 15383KB (283253KB) 3/19/04	<a href="#">42101_1999</a> 9812KB (173441KB) 3/17/04	<a href="#">42101_1998</a> 10031KB (176125KB) 3/17/04	<a href="#">42101_1997</a> 10136KB (177404KB) 3/17/04	<a href="#">42101_1996</a> 10102KB (176122KB) 3/17/04	<a href="#">42101_1995</a> 10171KB (175888KB) 3/17/04	<a href="#">42101_1994</a> 10774KB (185853KB) 3/17/04
Lead - PB - daily (12128) RD_501_12128_yyyy.ZIP	<a href="#">12128_2003</a> 47KB (637KB) 3/13/04	<a href="#">12128_2002</a> 65KB (872KB) 3/13/04	<a href="#">12128_2001</a> 70KB (926KB) 3/13/04	<a href="#">12128_2000</a> 66KB (919KB) 3/13/04	<a href="#">12128_1999</a> 72KB (996KB) 3/13/04	<a href="#">12128_1998</a> 85KB (1190KB) 3/13/04	<a href="#">12128_1997</a> 104KB (1484KB) 3/13/04	<a href="#">12128_1996</a> 120KB (1700KB) 3/13/04	<a href="#">12128_1995</a> 126KB (1774KB) 3/13/04	<a href="#">12128_1994</a> 136KB (1895KB) 3/13/04
Nitrogen Dioxide	<a href="#">42602_2003</a> 12708KB	<a href="#">42602_2002</a> 12051KB	<a href="#">42602_2001</a> 12795KB	<a href="#">42602_2000</a> 12712KB	<a href="#">42602_1999</a> 12271KB	<a href="#">42602_1998</a> 12228KB	<a href="#">42602_1997</a> 12708KB	<a href="#">42602_1996</a> 12712KB	<a href="#">42602_1995</a> 12540KB	<a href="#">42602_1994</a> 12767KB

- System Overview
- AQS Web Application
- Manuals and Guides
- Precision and Accuracy Data
- Frequently Asked Questions
- Memos
- User Registration
- Training
- AQS Conference
- Requesting AQS Data**
- Technical Forum
- Contacts
- Related Links**



# Data Warehousing

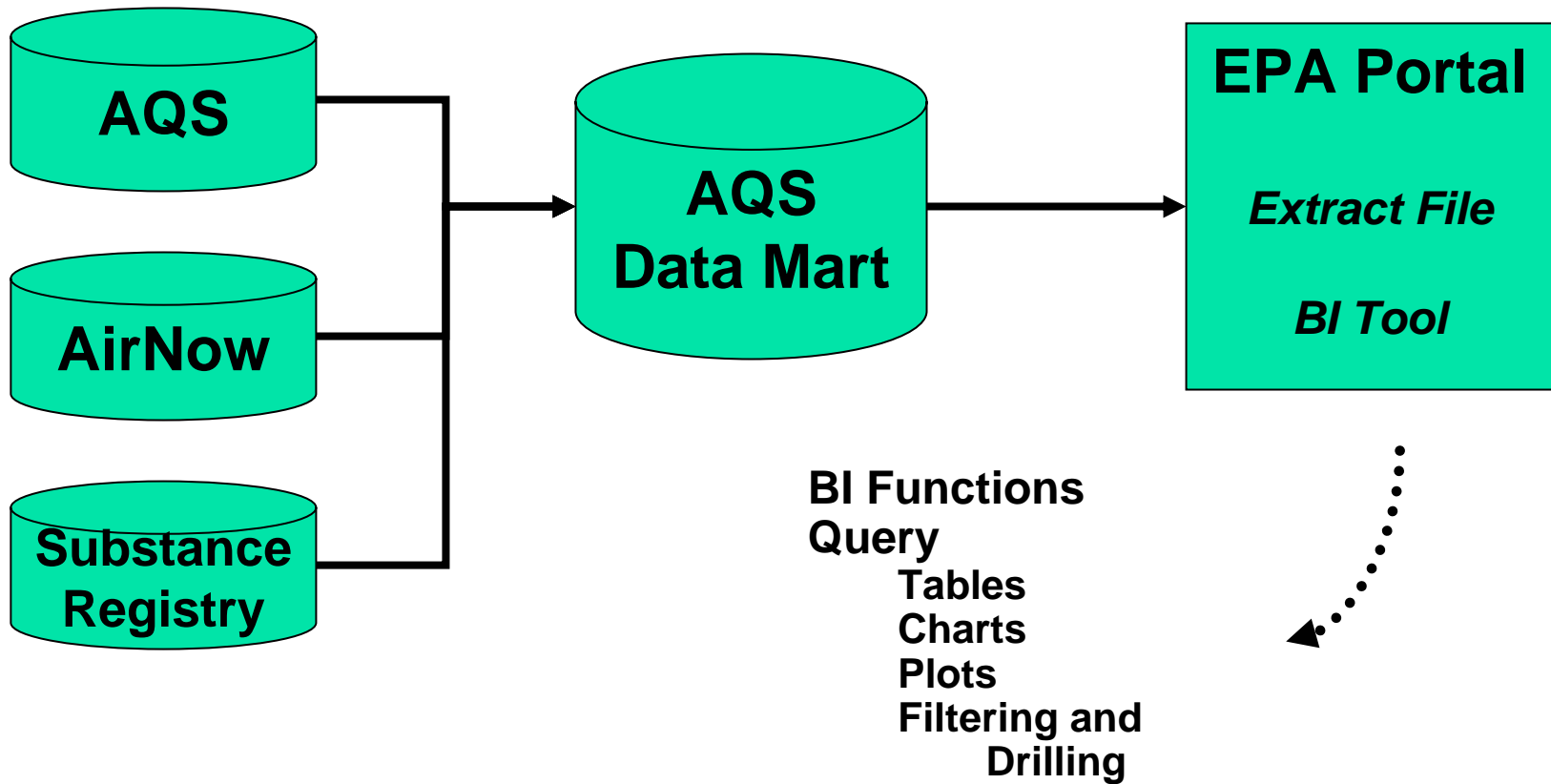
---

- Making data easier to query
  - Move (“replicate”) it to a different environment from the transaction system
    - Canyon >90% 5 out of 5 days last week
  - Reorganize it to make it faster to query
    - AQS optimized for input
  - Add other data useful for integration and analysis
    - Common definitions
    - Other program / media info
  - Connect “drag-and-drop” query tools to the front-end (“Portal”)
    - Example: Air Explorer, Business Intelligence



# AQS Data Mart Concept

---





# What's Covered

---

- All site, monitor, protocol, etc. data
- All raw and summary data
  - Excepting precision and accuracy
- 1980 – Present
- AirNow data until “same” submission comes into AQS
- Issues
  - Geographic related data quality at site level
    - May undertake a QA project soon
  - Descriptive data and caveats that would make the data more useable
    - “Field Guide to AQS Data”

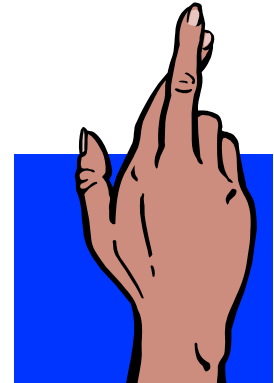


# Timing

---

- Project...

- We're working on this now
  - Project is in the design phase
- Demo system in September
- Production system in January



- Data...

- Nightly updates for data mart
- AQS data time lags



The logo graphic consists of three overlapping squares: a yellow one at the top left, a red one at the bottom left, and a blue one at the bottom right. A black crosshair is superimposed over the squares, with the vertical line passing through the center of the yellow and red squares, and the horizontal line passing through the center of the red and blue squares.

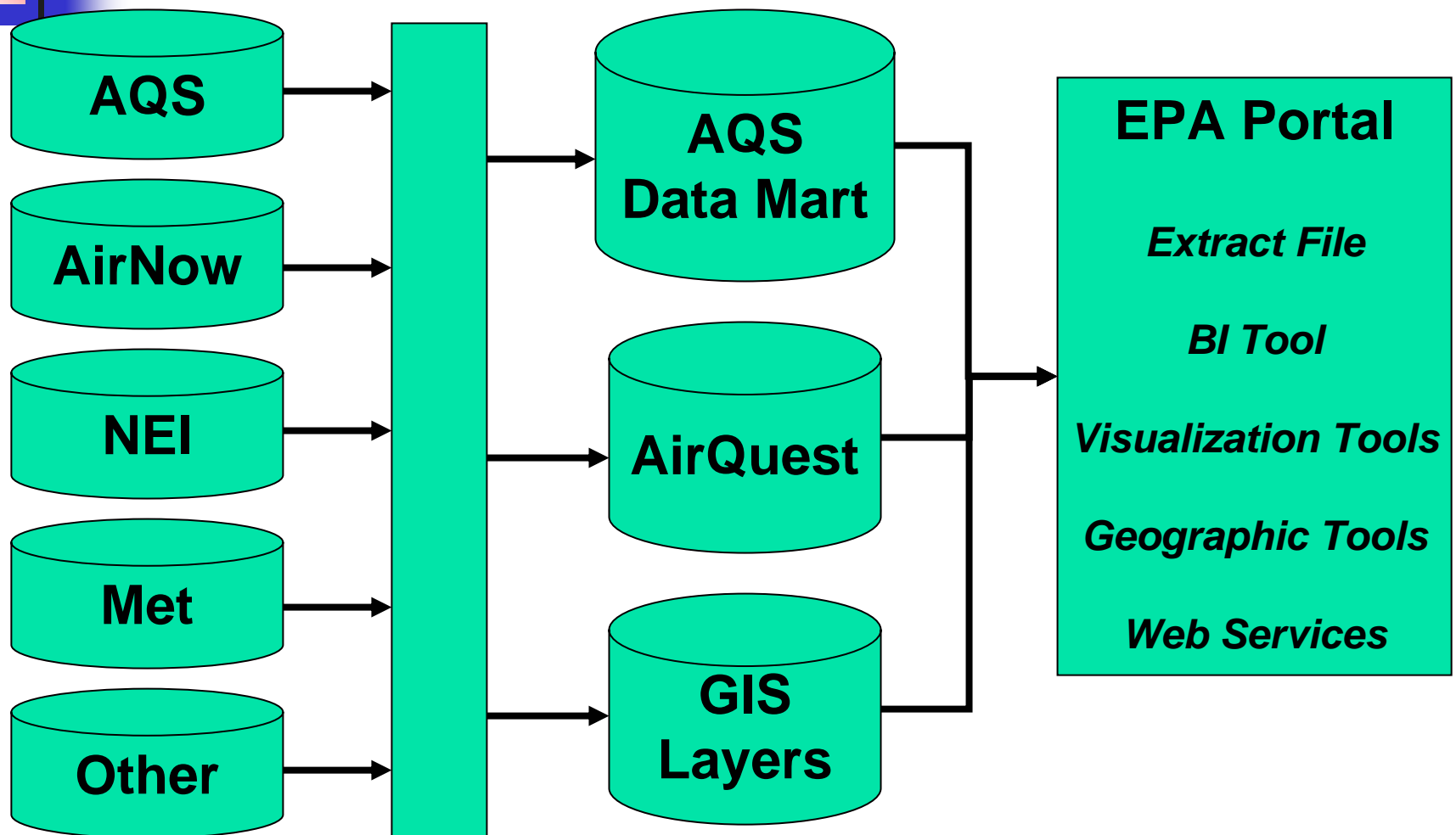
# AirQuest

---

- What:
  - Another data warehouse with the following data
    - AQS Criteria Summaries
    - NEI Criteria
    - Acid Rain CEM data
    - Acid Rain fuel and control data
    - Mercury stack tests
    - Coal pile / delivery samples
    - Emission factors
    - Meteorology
    - Demographics
    - More Coming soon: NEI Toxics, NATA results, PM2.5 Speciation from AQS, AirNow, gridded model results
  - Updated quarterly
  - Internal to EPA/OAQPS right now



# AQS Data Mart + AirQuest Concept





# Portal

---

- Portal
  - Initially users will be limited
  - Priority give to partners
    - You!
    - Researchers
  - Eventually opened up to anyone
  - Collection of tools and access to data
  - Customizable to the user's interest



# Wrap Up

---

- Tribal codes
- Data timeliness
- Data quality
- New technologies
  - XML
  - Warehousing
- “Connecting the dots”
- Thanks to OEI and AIEO for technical and financial support